

**An increase in marine heatwaves despite decreasing surface ocean
temperature variability**

**Tongtong Xu^{1*}, Matthew Newman^{1,2}, Antonietta Capotondi^{1,2}, Samantha Stevenson³,
and Emanuele Di Lorenzo⁴**

¹NOAA Physical Sciences Laboratory, Boulder, CO.

²CIRES, University of Colorado, Boulder, CO.

³Bren School of Environmental Science and Management, University of California, Santa Barbara, CA.

⁴Program in Ocean Science and Engineering, Georgia Institute of Technology, Atlanta, GA.

Submitted to Nature Communications

February, 2022

*Corresponding Author

Tongtong Xu (tongtong.xu@noaa.gov)

26 **Abstract (100-150 words)**

27 It is natural to assume that the increase in Marine Heatwaves (MHWs) - extremely warm
28 excursions in sea surface temperature (SST) – observed over 1958-2017 was caused by an
29 increase in SST variability. However, nonlinear changes in mean climate, if not accounted for
30 properly, could also cause apparent changes in SST variability. Here we use a large ensemble of
31 1958-2017 simulations from a Linear Inverse Model to show that the background warming trend,
32 as well as its nonlinear growth, was responsible for the observed increase in MHW events
33 worldwide, despite a significant overall reduction of SST variability in many oceanic regions
34 including the southern Indian, Atlantic, and Pacific oceans, the eastern Mediterranean, and the
35 Kuroshio-Oyashio Extension in the North Pacific. Identifying such regions, where a warming
36 mean state masks declining variability, is crucial for isolating regional climate change impacts
37 and thereby creating strategies to deal with future changes in ocean extremes.

Main

Marine heatwaves (MHWs) are events characterized by prolonged anomalously warm ocean surface or subsurface temperatures^{1, 2}, documented around the world^{3, 4}. Over the past several decades, these events have appeared to occur more often and become more extreme and longer-lasting^{5, 6}. The severe ecological and economic consequences of MHWs, including widespread mortality of marine species, adaptive reconfiguration of species ranges, and decline of farmed aquaculture production in commercial fisheries^{7, 8}, have motivated numerous studies aimed at understanding their atmospheric and oceanic origins^{4, 9, 10, 11}. Given current warming projections^{12, 13} and their potential impacts on future MHWs^{6, 12, 13, 14}, diagnosing observed MHW changes has gained increased urgency.

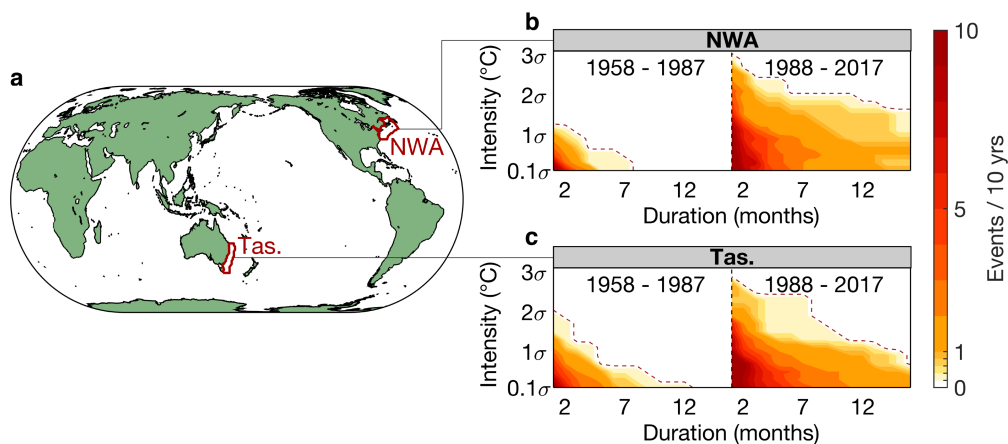


Figure 1. Examples of increased occurrence of more intense and longer lasting MHWs. (a) Global map with example regions marked by red outlines. Abbreviations: NWA, Northwest Atlantic; Tas., Tasman Sea. (b, c) MHW frequencies (events per decade) associated with events that reached a given threshold value of intensity and stayed at or above that level for at least a period of a given duration, i.e., Intensity-Duration-Frequency (IDF) plot at (b) NWA and (c) Tas., using the time series of monthly SST anomalies spatially averaged within each region (see

Methods), determined separately for the 30-yr periods of (left) 1958-1987 and (right) 1988-2017. Dashed line represents the contour level of 0.1 events per decade (i.e., 1 event/century).

The effects of climate change on MHWs may be quantified using measures of MHW intensity and duration^{15, 16}. Some examples are shown in Fig. 1, for two MHW hotspot regions, the Northwest Atlantic and the Tasman Sea³. For both, MHW occurrences during the years 1958-1987 and 1988-2017 are quantified by determining how often events reaching given intensity and duration thresholds occurred in each period, with the results displayed as Intensity-Duration-Frequency (IDF) plots (see ref. ¹⁵ and Methods). The frequency of MHW events notably increased between the two periods, e.g., events that rarely occurred during 1958-1987 occurred substantially more often during 1988-2017. This change was also found in many other regions^{3, 6, 17}, suggesting a near-global influence of climate change on MHWs.

These observed increases in extreme MHW events could simply be a consequence of increasing internal sea surface temperature (SST) variability, which by widening the probability distribution of ocean temperatures would make extreme warm temperatures more likely. Additionally, positive change in the mean climatology is generally understood to shift distributions to higher values, so that mean background warming could also result in an increase in events that reach or exceed some MHW threshold value, i.e., more occurrences and higher intensities of MHWs¹⁷. However, SST warming due to climatological change has been observed to be occurring at a *nonlinear* rate that has increased in the past few decades^{18, 19}. We will illustrate how this changing trend also acts to broaden the distribution, resulting in an *apparent* increase in variability and more extreme heatwaves. To date, most MHW studies have not differentiated between all these effects.

Separating the influences of mean-state and variability changes upon the observed increase in MHW events is key to guiding adaptation and mitigation management decisions responding to future MHW changes. In this paper, we conduct a global analysis of MHW events over the past 60 years that separates the long-term trend, crucially including its nonlinear change over time^{18, 19}, from the internal ocean variability, thereby avoiding conflation of the two and allowing for better estimates of the impact of climate change on both MHWs^{14, 20, 21} and overall SST variability²². We test the significance of these effects against large climate ensembles, or “alternative histories”, generated by a multivariate empirical dynamical model, specifically a Linear Inverse Model^{23, 24} (LIM; see Methods) that represents spatially-varying climate anomalies as a combination of linearly deterministic (i.e., predictable) dynamics plus an unpredictable nonlinear residual approximated by white noise. When run as a forecast model, the LIM makes seasonal-to-interannual predictions of surface ocean conditions, with forecast skill comparable to that of more sophisticated operational models^{25, 26 27}, and also diagnoses variations in potential predictability. When run as a climate model, the LIM generates climate realizations whose spatiotemporal evolution is statistically consistent with past observations, allowing their use for testing hypotheses including whether El Niño has significantly changed over the last several decades²⁸, the extent to which tropical Pacific decadal variability is a residual of El Niño events²⁹, and the impact of tropical and extratropical processes upon evolving Northeast Pacific MHW events¹⁶.

Motivated by these earlier results, we generate large LIM ensembles that realistically represent anomalous SST evolution observed over the past 60 years (1958-2017) to test the significance of changes in both SST mean and variance before and after 1987. These ensembles,

which we show capture the observed changes of MHW IDFs, are then used to evaluate relative impacts of the long-term trend and changes in internal variability upon MHW occurrences.

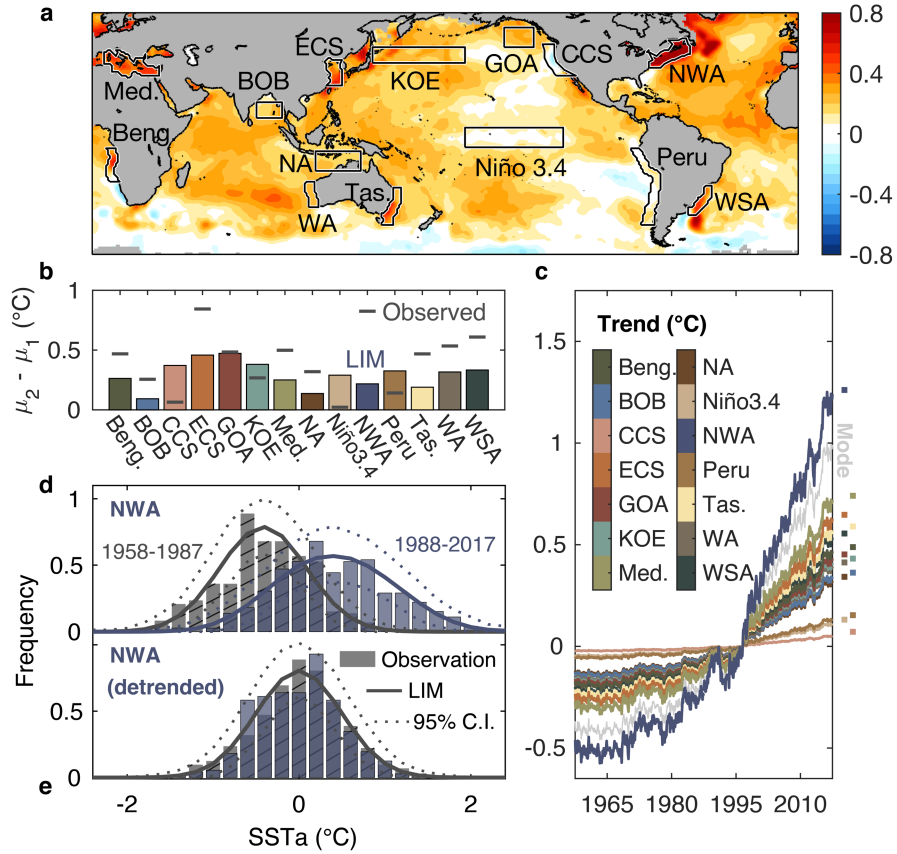


Figure 2. SST trends over the historical period. (a) Pattern of the trend mode over 1958-2017 (see Methods). Abbreviations: Beng., Benguela; BOB, Bay of Bengal; CCS, California Current System; ECS, East China Sea; GOA, Gulf of Alaska; KOE, Kuroshio-Oyashio Extension; Med., Mediterranean; NA, northern Australia; WA, Western Australia; WSA, Western South Atlantic. (b) Mean SST (μ) shift from 1958-1987 (μ_1) to 1988-2017 (μ_2) at each targeted region, calculated using the time series of monthly SST anomalies spatially averaged within each region. Horizontal lines are mean shifts determined from observed SSTa. Bars mark the edges of the 95% confidence interval of mean shifts obtained from a 3000-member LIM ensemble, where the

LIM was constructed from observations over the 1958-2017 period (the “LIM5817” ensemble; see Methods). (c) Observed trend of each targeted region, determined from the trend mode. Gray line is the time series of the trend mode. (d) Probability Distribution Functions (PDFs) of SSTa during 1958-1987 (gray) and 1988-2017 (blue) for the Northwest Atlantic region (NWA). Bars show the *observed SSTa* histograms and solid curves represent the ensemble mean PDFs for each period separately determined from the *trend+LIM5817* ensemble, with the dotted lines representing the 95% confidence intervals. (e) PDFs of detrended NWA SSTa during the two periods. Bars show the *detrended SSTa* histogram for the two periods, and solid and dotted curves represent the ensemble mean and the 95% confidence intervals determined from the *LIM5817 ensemble*.

Mean SST change and the observed nonlinear trend. We begin by identifying the pattern and amplitude of the observed long-term trend and estimating where it is statistically significant. Following previous studies^{23, 25, 26}, we use the least damped eigenmode^{30, 31} of the LIM’s dynamical operator to represent the trend component (see Methods). This trend pattern (Fig. 2a) appears generally robust, since similar patterns are obtained from other approaches, such as either computing the “mean shift” equal to the difference between the 1958-1987 and 1988-2017 means (Supplementary Fig. 1 and Fig. 2b) or fitting a piecewise linear trend at each grid point (see supplementary information); similar results have also been obtained in other studies^{22, 32}. The time series associated with the trend pattern (gray line in Fig. 2c) accelerates during the second half of the observational period, which is also seen in the evolution of regionally averaged SST anomalies (Fig. 2c).

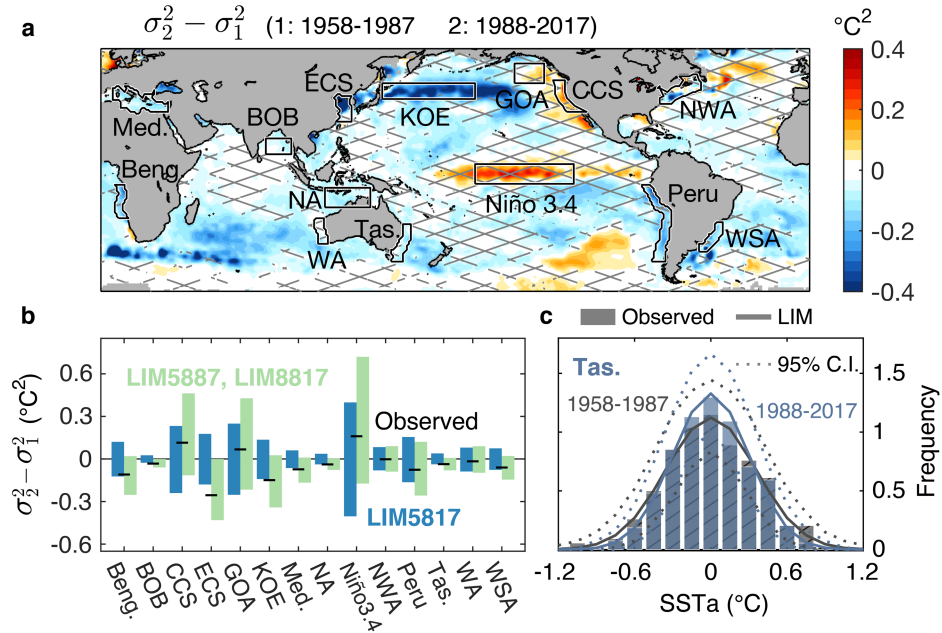
We tested the local statistical significance of the observed trend against a very large set of alternative histories based on the LIM’s realistic representation of the observed dynamical

system²⁷. After removing the least-damped eigenmode, the LIM is used to generate a 3000-member ensemble of 60-year-long different climate realizations of 1958-2017 internal variability (hereafter denoted LIM5817; see Methods). For each region, an observed mean shift was considered 95% significant only if it fell outside the 2.5% - 97.5% range of the 3000 mean shifts (shown by the bars in Fig. 2b) drawn from this LIM ensemble. Figure 2b shows that the regions of Benguela, Bay of Bengal, East China Sea, Gulf of Alaska, Mediterranean, northern Australia, Northwest Atlantic, Tasman Sea, Western Australia and Western South Atlantic all experienced significant mean shifts (see also Supplementary Fig. 1).

Impact of the observed nonlinear trend on variance. In many regions, observed historical trends have accelerated in the past few decades^{18, 19} (Fig. 2c), which as noted in the introduction may result in an apparent variance increase. This effect is perhaps most pronounced in the Northwest Atlantic region. The observed SSTa PDF (histogram bars in Fig. 2d) from the 1988-2017 period not only shifted to warmer values compared to the earlier 1958-1987 PDF, but was notably wider, which might be interpreted as clear evidence of an increase in variability. However, when we first remove the nonlinear trend (Fig. 2c) from Northwest Atlantic SSTa, we find detrended SSTa PDFs (histogram bars in Fig. 2e) with almost identical mean and variance for each 30-yr period.

We can use the LIM ensembles to make sense of these results. First, note that the PDFs determined from the LIM5817 ensembles fit both 1958-1987 and 1988-2017 detrended SSTa histograms (solid and dotted lines in Fig. 2e), indicating that the detrended two periods are statistically indistinguishable. Next, the observed trend component is added back to each LIM5817 ensemble member, so that the resulting “trend+LIM5817” ensemble now represents

159 both the trend and the variability during 1958-2017. PDFs are then determined from the
 160 trend+LIM5817 ensembles, separately for the first half (corresponding to 1958-1987) and the
 161 second half (corresponding to 1988-2017) of the 60-yr period. By construction, any difference
 162 between these two simulated PDFs is only due to the externally-forced trend, since the
 163 underlying internal variability is unchanged (i.e., it is generated by a single LIM, as in lines of
 164 Fig. 2e). Figure 2d shows that the simulated PDFs match the observed PDFs, notably capturing
 165 the observed PDF change between the two periods, suggesting that the accelerating trend was
 166 responsible not only for the overall shift to warmer values but also for the widening of the PDF.
 167 A similar result is seen in other regions with strong nonlinear trends (e.g., the Mediterranean
 168 region; see Supplementary Fig. 2), while the effect is weaker in regions with weaker trends (see
 169 Supplementary Fig. 2 for other regions).



170
 171 **Figure 3. Change in the variance of detrended SSTa over the historical period.** (a) Map of
 172 local variance changes from 1958-1987 to 1988-2017. Shading shows the difference in observed
 173 variance between the two periods, in units of $(^{\circ}\text{C})^2$. Regions that are *not* cross-hatched indicate

locations where the observed variance changes are significant (40% of the ocean domain), i.e., outside the 95% range of variance changes within the LIM5817 ensemble. (b) SST variance (σ^2) change from 1958-1987 (σ_1^2) to 1988-2017 (σ_2^2) at each targeted region, calculated using the time series of monthly SST anomalies spatially averaged within each region. Horizontal lines are variance changes determined from observed detrended SSTa. Blue bars show the 95% range of variance changes between the two periods obtained from the 3000-member LIM5817 ensemble; green bars show the 95% range of variance changes between the two periods but obtained from LIMs constructed over the two periods separately (“LIM5887” and “LIM8817”; see Methods). (c) PDFs of detrended SSTa during 1958-1987 (gray) and 1988-2017 (blue) at Tasman Sea region (Tas.). Bars are PDFs of observed detrended SSTa. Solid curves and dotted lines are the ensemble mean and 95% confidence interval, respectively, of LIM5887 and LIM8817’s PDFs.

Steady or slightly decreasing internal SST variability. While the observed intensifying trend has been shown to contribute to an apparent increase in variance, it remains possible that the underlying internal variance could also have changed. To assess such potential changes, we first determine “detrended variance” from the observed detrended SSTa. We then estimate the difference between the detrended variance for the two periods 1958-1987 and 1988-2017; results are shown globally in Fig. 3a (shading) and regionally in Fig. 3b (horizontal lines). While detrended variance increased in a few regions, notably within the Niño 3.4 and the Northeast Pacific (Fig. 3b), over most of the globe it decreased (Fig. 3a), particularly within the East China Sea and Kuroshio-Oyashio Extension (Fig. 3b). Decreases in detrended variance are also reflected in narrowed PDFs from 1958-1987 to 1988-2017, such as in the Tasman Sea region (histogram bars in Fig. 3c).

Next, we assess the significance of the observed changes in detrended variance, by asking whether they could have occurred by chance within a statistically stationary system. This question is answered by comparing these observed changes against the potential changes in variance seen in the LIM5817 ensemble (see Methods); 95% significance levels are indicated by cross-hatching in Fig. 3a and blue bars in Fig. 3b. Interestingly, while many of the notable observed variance decreases are significant (e.g., in the Bay of Bengal, East China Sea, Kuroshio-Oyashio Extension, Mediterranean, northern Australia and Tasman Sea regions), the variance increases are not.

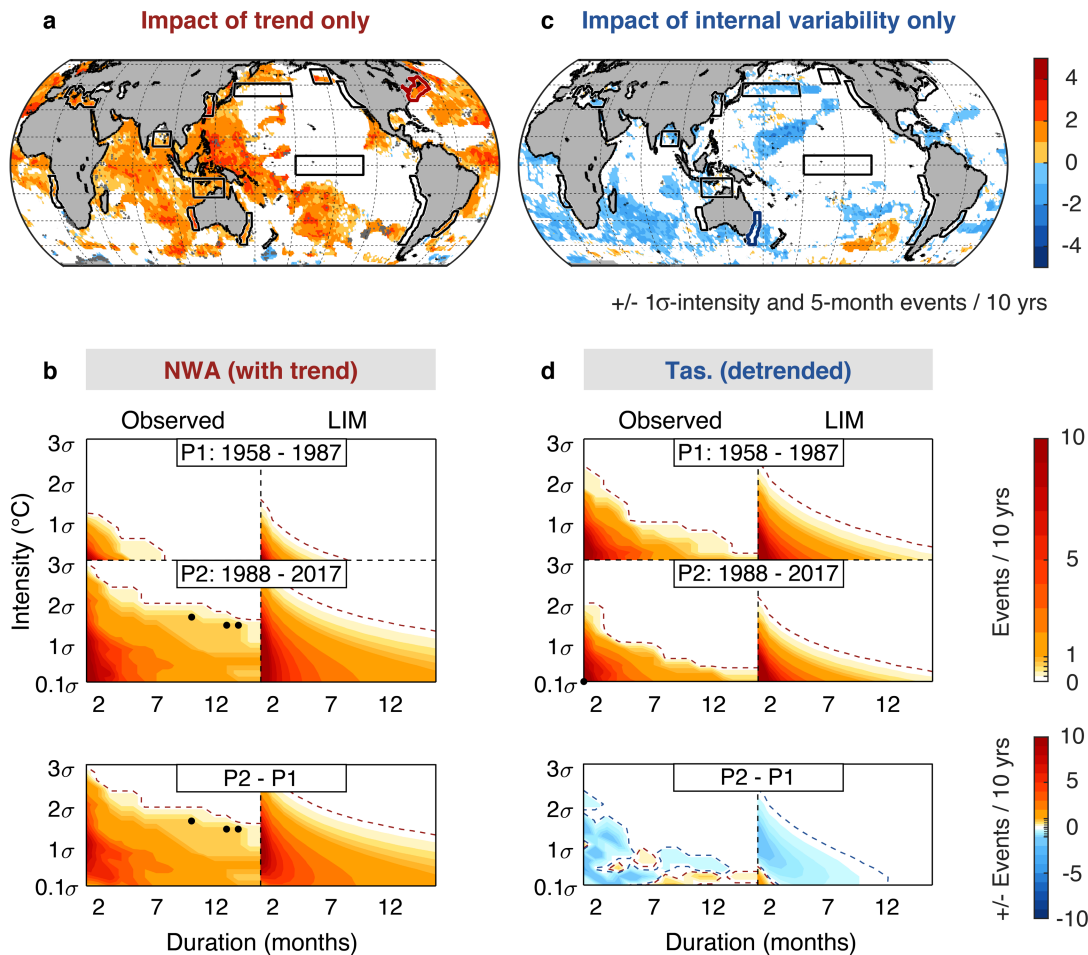


Figure 4. Impact of changes in either the trend or internal (detrended) variability on MHW occurrences. (a) Global change in the frequency (events per decade) of MHW events, here

defined as those reaching an amplitude (intensity) of at least 1σ and thereafter persisting for at least 5 months (duration), between 1958-1987 and 1988-2017 (see Methods), derived from *observed SSTa*, displayed only for locations where the mean shifts were 95% significant in Supplementary Fig. 1. Gray dots mark where the observed MHW frequencies are 95% significantly different from LIM frequencies (see Methods). (b) IDF plot for the Northwest Atlantic region (NWA; region marked by red outline in panel a), determined from *observed SSTa* (left column) and the “*trend+LIM5817*” ensemble (right column). The first row shows the IDF of 1958-1987 (P1), the second row shows the IDF of 1988-2017 (P2), and third row is the difference of the IDFs between the two periods (P2 - P1). Dashed lines represent the contour level of 0.1 events (red) or -0.1 events (blue) per decade. Black dots mark observed values that are 95% significantly different from the LIM (i.e., outside 95% of the LIM ensemble); here they represent below 5% of the IDF values and so are not field significant. (c) Global change in the frequency of MHW events, defined as in (a), except the statistics were derived from *detrended SSTa* and are displayed only for locations where detrended variance changes were 95% significant in Fig. 3a. (d) IDF plot for the same periods as in (b), but for the Tasman Sea region (Tas.; region marked by blue outline in panel c) and determined from *detrended SSTa* (left column) and the *LIM5887 and LIM8817 ensembles* (right column), respectively.

Separating the effects of changes in trend and variability upon MHW occurrence. We next analyze how the changes in the observed trend and internal variability each (separately) impact MHW occurrence, as captured by the IDF plots that measure the frequencies of MHWs with differing durations and intensities. In both cases, we compare observed IDF plots for the 1958-1987 and 1988-2017 periods, and their differences, with matching IDF plots determined from LIM large ensembles. The results of this analysis are shown in Fig. 4.

First, we return to the Northwest Atlantic to show how the trend alone can impact MHW IDF plots. The observed IDF plots for the two periods are shown in the left column of Fig. 4b, repeated from Fig. 1b, with their differences shown in the third row. The corresponding IDF plots from the trend+LIM5817 ensembles (right column of Fig. 4b) almost entirely reproduce the observational results. These figures show that extreme Northwest Atlantic MHW events became more frequent for a wide range of intensities and durations. The strong correspondence between the observed and trend+LIM5817 ensemble IDF difference plots demonstrates that these changes can be almost entirely attributed to the effect of the trend.

The trend similarly impacts the frequency of MHW events for the other regions examined in this study, with the intensity of this effect related to the strength of the warming (Supplementary Fig. 3). To capture the global extent of the trend effect, we repeat the analysis of Fig. 4b at each ocean grid location, i.e., each grid point has its own set of IDFs. To condense this analysis into a single figure, we show results using one representative definition of MHW events, those that reach at least 1σ intensity and persist for at least 5-month duration (Fig. 4a), chosen to match previous studies of Northeast Pacific MHWs^{16, 33}. Choosing other pairs of thresholds to define MHW events yields a qualitatively similar picture (see Supplementary Fig. 5). Note that in Fig. 4a, results are displayed only for locations where the mean shifts are significant (as indicated by Supplementary Fig. 1). Finally, we test where these values are significantly different from the ensemble statistics derived from trend+LIM5817; these regions are also indicated in the figure (gray dots). Overall, the change in frequency of observed MHWs, associated with both the threshold pair chosen for Fig. 4a and various other intensity and duration pairs, are generally reproduced by the ensemble statistics derived from trend+LIM5817 (see also Supplementary Fig.

4b and 5ab). This indicates that the trend induced mean changes is responsible for these observed MHW changes.

We next evaluate how changes to detrended variability alone impacts the occurrence of different MHW events. In this case, we return to the Tasman Sea region, but instead of displaying the observed IDF plots for the two periods derived from the *total* SSTa, as was shown in Fig. 1c, we instead show the IDF plots from *detrended* SSTa in the left column of Fig. 4d, with the differences in the third row. Again, we could compare the observed change in the IDF plots to the change in the LIM5817 ensembles. However, since we have already determined that this region underwent a significant variance decrease (Fig. 3b), we instead construct *two* distinct LIMs -- one from the detrended anomalies of each 30-year period (referred to as LIM5887 and LIM8817; see Methods) -- and then use each to generate 3000-member 30-yr ensembles that can represent possible changes in dynamics and hence changed variability (as shown in the green bars of Fig. 3b). In this case, the resulting simulated IDF plots for the two periods in the Tasman Sea region (right column of Fig. 4d) still match the observed IDFs. Note that both sets of IDF plots show a large decrease of occurrences over a wide intensity range of short-lived MHWs, a feature also seen in IDFs for other regions that undergo a decrease in detrended variance (Supplementary Fig. 3).

We again analyze the change in detrended variability effect on MHWs at all locations, using the same representative definition for MHWs of Fig. 4a, with results displayed in Fig. 4c but only at locations where the changes of detrended variances are significant (as indicated by Fig. 3a). Overall, we find that if we had only considered *detrended* observed SSTa, MHW occurrences would have *decreased* in many regions of the globe, which is generally reproduced by the MHW frequency changes from the LIM5887 to the LIM8817 ensembles (Supplementary

Fig. 4c and 5cd). Note that there appears to be no obvious pattern in Fig. 4c, nor does it bear a relationship to the trend pattern in Fig. 4a.

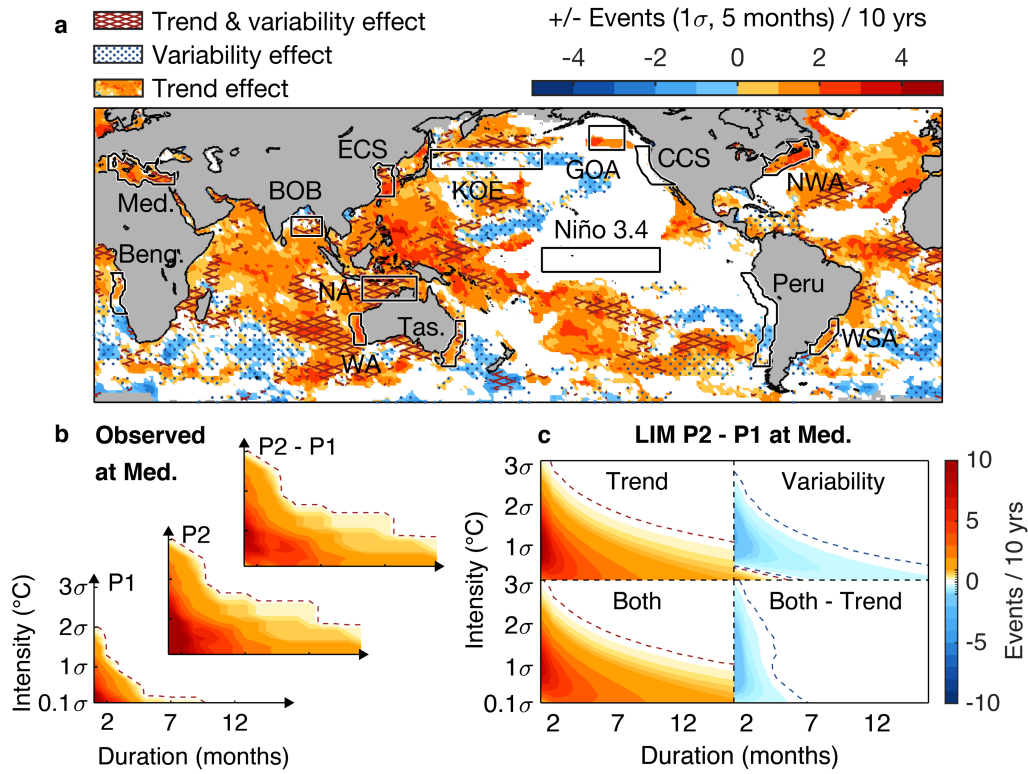


Figure 5. Combined effect of trend and detrended variability on MHW occurrences. (a) Observed change in the frequency of MHW events, defined as in Fig. 4a, in units of change in events per decade. Regions where MHW change is due to both trend and internal variability are indicated by cross-hatching; regions where MHW change is due to the detrended variability only are indicated by dots; regions where MHW change is due to the trend only are not stippled; regions that are white do not have significant changes in either trend or variability. (b) Observed IDF plot for the same periods as in Fig. 4b, but for the Mediterranean region (Med.). (c) Simulated IDF difference plot at Med., derived from (top-left) trend+LIM5817, (top-right) LIM5887 and LIM8817, (bottom-left) trend+LIM5887 and trend+LIM8817 (see Methods), and (bottom-right) differences between the two left panels.

Collective trend and variability effect on MHWs. Finally, after examining how changes to the trend and the detrended variability separately impact MHW occurrences, we analyze how they collectively contribute to MHW changes. Figure 5a summarizes our results, showing both the observed change in MHW frequency (shading), and which effect(s) is (are) responsible for those changes, at each location worldwide. We again define MHW events with the threshold pair used for Figs. 4a and c, and as before our main results are not qualitatively sensitive to this choice. In Fig. 5a, regions where MHW frequency is significantly impacted by *both* changes in the trend and in detrended variance are indicated by cross-hatching; note that these are locations shaded in both Fig. 4a and c. Regions with significant changes due only to detrended variance are indicated by dotted areas, while regions of only significant trend effect have neither hatching nor dotting. The extensive white regions have no significant changes. Overall, the global map reveals that, collectively accounting for the trend and the internal variability effects, large regions with significant trend effect have increased MHW occurrences (e.g., Indian Ocean), whereas decreased occurrences are found in somewhat smaller regions with significant variance changes (e.g., Kuroshio-Oyashio Extension).

For the most part, in the regions that experienced both significant trend and internal variability changes, the occurrences of MHWs have increased. For example, the Mediterranean experienced both a significant decrease in detrended variance (Fig. 3b) and a significant increase in the mean (Fig. 2b). To diagnose the net impact of these two effects, we compare the observed Mediterranean IDF plots (Fig. 5b) to the LIM IDFs (Fig. 5c), where we now use trend+LIM5887 and trend+LIM8817 ensembles that capture both the trend and change in variance. We find that the Mediterranean would have experienced a decreased frequency of MHWs solely under the

effect of internal variability (top-right of Fig. 5c) or would have experienced stronger increases of MHWs solely under the effect of the trend (top-left). Consequently, the combined effect leads to a *lessened* increase on the MHW statistics (bottom-left of Fig. 5c) and predominantly affects the more intense occurrences rather than their persistence (bottom-right).

Discussion. Recent research has identified many regions worldwide where MHWs appear to be recently occurring more frequently and with increased severity. However, it is striking that in almost all these MHW hotspots, apart from those located in the Northeast Pacific, SST variance – relative to the background trend – has either stayed relatively constant or has even significantly *decreased* over the past 60 years. The reasons for the SST variability decrease remain unclear; note that our LIM analysis shows only that there has been a significant change in variance but not its cause. With this decreasing variance, the occurrences of climate extremes, such as MHWs, should have also decreased. Yet we instead found increasing occurrences of MHWs over a wide range of intensities and durations in these MHW hotspots. These increases are entirely attributable to the observed historical trend, which has been accelerating over time, shifting the mean background climatology to become increasingly warmer at an increasing speed. This observed intensifying trend has therefore led to both a mean warming effect and even, in some locations, an apparent increase of the variance, both shifting the probability distribution to warmer values and widening it. The impact on MHW frequency of the observed warming trend alone was generally strong enough to offset the effect driven by the reduction of SST internal variability wherever it occurred.

Finally, although the historical period has experienced a nonlinearly growing trend and an overall decrease of internal variability, it remains to be assessed whether such changes will

continue in the remaining portion of the 21st century, and what their impacts on the occurrence of future MHWs – and other marine extremes – will be. Our LIM framework provides a clear path for such future assessments.

Methods

Sea surface temperature (SST) and sea surface height (SSH) data. Monthly SSTs from the Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST)³⁴ and monthly SSHs from the European Centre for Medium-Range Weather Forecasts (ECMWF) Ocean Reanalysis System 4 (ORAS4)³⁵, during the period 1958-2017, were analyzed. The global domain extended from 60°S to 64°N, with 1°×1° spatial resolution. The climatological annual cycle computed from the full length of the historical record at each grid point was removed to obtain the SST anomaly (SSTa) and SSH anomaly (SSHa).

Linear Inverse Model (LIM). The time evolution of a climate state \mathbf{x} may often be approximated by the stochastically forced linear dynamical system,

$$\frac{d\mathbf{x}}{dt} = \mathbf{L}\mathbf{x} + \boldsymbol{\xi} \quad (1)$$

where $\mathbf{x}(t)$ is the climate state, \mathbf{L} is a linear dynamical operator, $\boldsymbol{\xi}$ is a vector of temporally white noise that may have spatial structure (determined from a balance condition derived from (1)), and t is time. Determining (1) from observed covariances results in a Linear Inverse Model (LIM24). LIMs are typically low-order models, where the state vector is expressed in reduced Empirical Orthogonal Function (EOF) spaces. In this paper, $\mathbf{x}(t)$ represents the leading Principal Components (PCs) of SSTa (29 PCs) and SSHa (22 PCs), where the EOFs for each field were separately determined, explaining 74.9% and 73.7% of each corresponding field's total variance.

The lag-covariance used to determine the LIM operators was computed using a training lag of $\tau_0 = 1$ month. We also tested to make sure the results were not sensitive to this choice, as is generally done when constructing a LIM. See 16, 26, 27, 36 for other details concerning the LIM and its construction.

Process of identifying the trend from the linear dynamical operator. Several studies^{30, 31} have shown how the externally forced trend is captured by the least damped eigenmode of \mathbf{L} . To identify the trend, we performed the eigenanalysis on \mathbf{L} ; that is,

$$\mathbf{L}\mathbf{U} = \mathbf{U}\mathbf{\Lambda} \quad (2)$$

where \mathbf{U} is the matrix of eigenvectors and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues (λ_i). \mathbf{V} , the eigenvectors of \mathbf{L} 's adjoint, is simply determined by $\mathbf{V}^H = \mathbf{U}^{-1}$, such that $\mathbf{L}^H\mathbf{V} = \mathbf{V}\mathbf{\Lambda}^*$, where $()^H$ is the conjugate transpose and $()^*$ is the conjugate. The least damped mode has the longest decay time; that is, it is associated with the eigenvalue λ_i with the largest value of $|1/\text{Re}(\lambda_i)|$. The spatial pattern (Fig. 2a) of the least damped mode (\mathbf{u}_i) is obtained as the i -th column of \mathbf{U} , with its time series (gray line in Fig. 2c) obtained by $\mathbf{v}_i^H \mathbf{x}(t)$, where \mathbf{v}_i is the i -th column of \mathbf{V} . Thus, we identify the trend component, as the projection of $\mathbf{x}_{\text{TR}}(t) = \mathbf{u}_i \mathbf{v}_i^H \mathbf{x}(t)$.

Note that $\mathbf{x}_{\text{TR}}(t)$ is only derived once based on the full length of the observed record we examined. The presented spatial pattern in Fig. 2a is normalized by its spatial maximum. The presented time series in Fig. 2c is normalized by its temporal maximum. We obtained generally similar results by instead removing the piecewise linear trend; see supplementary information for details of this alternative approach.

LIM climate simulations. LIM simulations may be generated by integrating Equation (1) forward in time, driven by white noise forcing with observationally constrained spatial structure²³. In addition, the remaining variance (i.e., contained in the unresolved PCs that were not used in LIM construction), denoted as \mathbf{x}_U , is approximated as purely white noise,

$$\frac{d\mathbf{x}_U}{dt} = \xi_U \quad (3)$$

where the time-varying white noise is simply approximated by randomly energizing the amplitude of the PCs. See ³⁷ for details of incorporating the unresolved PCs to construct an untruncated LIM.

In this study, we constructed the LIM by first determining the linear dynamical operator using the observational record over the entire 60-year period. We conduct the eigenanalysis on the dynamical operator, obtain the least damped eigenmode as the LIM trend, and remove the trend component from the observed anomalies. The detrended anomalies are then used to construct the LIM for generating 3000-member ensemble of 60-year-long realizations, representing the detrended dynamical system of 1958-2017 years, i.e., the LIM5817 ensemble. Since the ensemble is generated using a dynamical operator that does not discriminate between the 1958-1987 and 1988-2017 periods, differences between the first and second halves of each 60-year periods can only arise from the system noise, i.e., the internal variability of the two periods is unchanged.

To represent the full dynamical system, we add the trend component back to each LIM5817 ensemble member, i.e., trend+LIM5817 ensemble. By construction, the statistical difference between the two 30-year periods of the trend+LIM5817 ensemble is only due to the externally-forced trend.

We also construct two *new* LIMs to represent each 30-year period, i.e., the LIMs that do not see the two periods equivalently. One LIM is constructed using the detrended anomalies of the 1958-1987 period, the generated ensemble therefore provides the detrended realizations of the 1958-1987 period, i.e., LIM5887 ensemble. Similarly, we construct the other new LIM using the detrended anomalies during 1988-2017, hence denoting the generated ensemble as LIM8817 ensemble. The two ensembles represent the detrended dynamical system of each the 30-year period, and any possible difference between the two ensembles reflects the observed changes of the internal variability during the two periods. Note that LIM ensembles of this study are untruncated LIM (see ³⁷ for details).

MHW frequency. Frequency is determined by calculating the number of events that exceed (\geq) a given intensity for a period longer than (\geq) a given duration, divided by the total number of years in the observational record, in units of events per 10 years. The IDF plot is derived from calculating the frequency for each intensity and duration threshold pair, including intensities from 0.1σ to 3.1σ and durations from 1 to 16 months. For the spatial maps of frequency differences, we first determine the frequency, for each 30 yr period and at each location, for MHW events defined by the 1σ °C intensity and 5 months duration threshold pair, as in a previous study¹⁶. Then the frequencies of the two periods are differenced for these global maps.

Significance tests. Multiple significance tests are carried out in this study, including testing the significance of the mean shifts, the variance changes, and whether the LIM reproduces the observed SST records and the MHWs. In general, the process is as follows: we obtain the observed value and compare that with the range of simulated values provided by the LIM climate

ensembles. If the observed value is within (outside) the 95% LIM ranges, we determine that as insignificant (significant). In detail,

(a) Significance of mean shifts and variance changes. We obtain the observed mean shifts by subtracting the mean of 1958-1987 from the mean of 1988-2017, in the input observational dataset. For each 60-yr LIM5817 ensemble, i.e., alternative 1958-2017 realization, we then compute its simulated mean shift. Since we constructed a 3000-member LIM5817 ensemble, we then have 3000 mean changes at each grid point; that is 3000 potential changes which can be compared to that observed. From these 3000 changes, we obtain the 2.5% and the 97.5% values, hence the 95% confidence intervals. These 95% ranges are compared to the observed mean shifts to determine its significance. Results are shown in Fig. 2b and Supplementary Fig. 1.

A similar process is carried out for testing the significance of the variance changes (Fig. 3a and blue bars of Fig. 3b). We obtain simulated variance changes by subtracting the variance of the first 30 year from the second in each of the LIM5817 ensemble and compare their 95% ranges to the observed detrended variance changes.

(b) Reproduction of variance changes. Since large regions of the global ocean have experienced significant decreases in the internal variability (shown by Fig. 3a and blue bars of Fig. 3b), we construct the two distinct LIMs, driven by the internal variability of 1958-1987 and 1988-2017, respectively (LIM5887 and LIM8817). These ensembles therefore represent the internal variability of the two periods. Subsequently, their variance differences should reflect the observed variance differences. To verify that is the case, we recomputed variance changes between the members of LIM5887 and LIM8817

ensembles. Their regionally averaged values are shown in the green bars of Fig. 3b, which provide a reasonable range for the observed variance changes.

(c) Reproduction of probability distribution. To show that our LIM ensembles realistically reproduce the observed record, we compute the PDFs of each LIM ensemble member and compare them to the observed PDFs. This is first carried out in the trend+LIM5817 ensemble, to validate that the ensemble is a realistic representation of the observed trend. That is, for each 60-yr trend+LIM5817 ensemble member, we compute the PDFs of the simulated 1958-1987 and the PDFs of the simulated 1988-2017. Given the 3000-member trend+LIM5817 ensemble, we have therefore 3000 PDFs for each period. We then obtain the mean, the 2.5% and the 97.5% values of each bin of the PDFs. These are the ensemble mean PDFs and their 95% confidence interval. If the ensemble-mean PDFs overall overlap with the observed PDFs, within the 95% confidence interval, we consider our LIM ensemble able to track the observations. Results are shown in Fig. 2d, and in Supplementary Fig. 2.

A similar process is carried out on the LIM5887 and LIM8817 ensembles, to validate that they realistically capture the internal variability of each 30-yr period. One example is shown by the simulated Tasman Sea PDFs compared to the observed (Fig. 3c). Our result shows the ensemble mean PDFs becoming narrower from 1958-1987 to 1988-2017, consistent with the observed PDF changes. Other regions also consistently show observed PDFs captured by these two LIMs (Supplementary Fig. 2).

(d) Reproduction of MHWs. To show that our LIM ensembles reproduce the MHW statistics of various intensities and durations, we first compute the observed IDF plots of the two 30-yr period and their differences at each grid location. We then derive the simulated IDF

plots and their differences from each LIM ensemble member, i.e., resulting in 3000 IDF difference plots at each grid point. These therefore give us the local 2.5% and the 97.5% IDF difference plots. We first focus on a representative pair of thresholds – 1σ intensity and 5-month duration – and check whether the observed frequency is within or outside the 95% range of LIM frequency. This is analysed at each grid point, using the trend+LIM5817 ensemble, with results shown by the gray dots in Fig. 4a, and using the LIM5887 and LIM8817 ensemble, shown in Fig. 4c. We next conduct the analysis to several other pairs of thresholds to show how our results are not sensitive to the choice of the representative thresholds (Supplementary Fig. 5).

REFERENCES

1. Hobday AJ, Alexander LV, Perkins SE, Smale DA, Straub SC, Oliver ECJ, *et al.* A hierarchical approach to defining marine heatwaves. *Progress in Oceanography* 2016, **141**: 227-238.
2. Hobday AJ, Oliver ECJ, Sen Gupta A, Benthuyssen JA, Burrows MT, Donat MG, *et al.* Categorizing and Naming MARINE HEATWAVES. *Oceanography* 2018, **31**(2): 162-173.
3. Oliver ECJ, Benthuyssen JA, Darmaraki S, Donat MG, Hobday AJ, Holbrook NJ, *et al.* Marine Heatwaves. *Annual Review of Marine Science* 2021, **13**(1): null.
4. Holbrook NJ, Scannell HA, Sen Gupta A, Benthuyssen JA, Feng M, Oliver ECJ, *et al.* A global assessment of marine heatwaves and their drivers. *Nature Communications* 2019, **10**: 13.

492

493 5. Oliver ECJ, Donat MG, Burrows MT, Moore PJ, Smale DA, Alexander LV, *et al.* Longer
494 and more frequent marine heatwaves over the past century. *Nature Communications* 2018, **9**: 12.

495

496 6. Laufkötter C, Zscheischler J, Frölicher TL. High-impact marine heatwaves attributable to
497 human-induced global warming. *Science* 2020, **369**(6511): 1621.

498

499 7. Holbrook NJ, Sen Gupta A, Oliver ECJ, Hobday AJ, Benthuisen JA, Scannell HA, *et al.*
500 Keeping pace with marine heatwaves. *Nature Reviews Earth & Environment* 2020.

501

502 8. Smale DA, Wernberg T, Oliver ECJ, Thomsen M, Harvey BP, Straub SC, *et al.* Marine
503 heatwaves threaten global biodiversity and the provision of ecosystem services. *Nat Clim*
504 *Change* 2019, **9**(4): 306-+.

505

506 9. Di Lorenzo E, Mantua N. Multi-year persistence of the 2014/15 North Pacific marine
507 heatwave. *Nat Clim Change* 2016, **6**(11): 1042-+.

508

509 10. Bond NA, Cronin MF, Freeland H, Mantua N. Causes and impacts of the 2014 warm
510 anomaly in the NE Pacific. *Geophysical Research Letters* 2015, **42**(9): 3414-3420.

511

512 11. Amaya DJ, Alexander MA, Capotondi A, Deser C, Karnauskas KB, Miller AJ, *et al.* Are
513 Long-Term Changes in Mixed Layer Depth Influencing North Pacific Marine Heatwaves?
514 *Bulletin of the American Meteorological Society* 2021, **102**(1): S59-S66.

515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536

12. Frolicher TL, Fischer EM, Gruber N. Marine heatwaves under global warming. *Nature* 2018, **560**(7718): 360-+.

13. Misra R, Serazin G, Meissner KJ, Sen Gupta A. Projected Changes to Australian Marine Heatwaves. *Geophysical Research Letters* 2021, **48**(7).

14. Shi H, Garcia-Reyes M, Jacox MG, Rykaczewski RR, Black BA, Bograd SJ, *et al.* Co-occurrence of California Drought and Northeast Pacific Marine Heatwaves Under Climate Change. *Geophysical Research Letters* 2021, **48**(17).

15. Scannell HA, Pershing AJ, Alexander MA, Thomas AC, Mills KE. Frequency of marine heatwaves in the North Atlantic and North Pacific since 1950. *Geophysical Research Letters* 2016, **43**(5): 2069-2076.

16. Xu T, Newman M, Capotondi A, Di Lorenzo E. The Continuum of Northeast Pacific Marine Heatwaves and Their Relationship to the Tropical Pacific. *Geophysical Research Letters* 2021, **48**(2): 2020GL090661.

17. Oliver ECJ. Mean warming not variability drives marine heatwave trends. *Climate Dynamics* 2019, **53**(3-4): 1653-1659.

537 18. Hartmann DL, Tank AMK, Rusticucci M, Alexander LV, Brönnimann S, Charabi YAR,
538 *et al.* Observations: atmosphere and surface. *Climate change 2013 the physical science basis:*
539 *Working group I contribution to the fifth assessment report of the intergovernmental panel on*
540 *climate change*. Cambridge University Press, 2013, pp 159-254.

541

542 19. Seidel DJ, Lanzante JR. An assessment of three alternatives to linear trends for
543 characterizing global atmospheric temperature changes. *Journal of Geophysical Research-*
544 *Atmospheres* 2004, **109**(D14): D14108.

545

546 20. Jacox MG. OCEAN SCIENCE Marine heatwaves in a changing climate. *Nature* 2019,
547 **571**(7766): 485-487.

548

549 21. Jacox MG, Alexander MA, Bograd SJ, Scott JD. Thermal displacement by marine
550 heatwaves. *Nature* 2020, **584**(7819): 82-86.

551

552 22. Alexander MA, Scott JD, Friedland KD, Mills KE, Nye JA, Pershing AJ, *et al.* Projected
553 sea surface temperatures over the 21st century: Changes in the mean, variability and extremes for
554 large marine ecosystem regions of Northern Oceans. *Elementa-Sci Anthropol* 2018, **6**.

555

556 23. Penland C, Matrosova L. A Balance Condition for Stochastic Numerical-Models with
557 Application to the El-Nino-Southern Oscillation. *Journal of Climate* 1994, **7**(9): 1352-1372.

558

24. Penland C, Sardeshmukh PD. The Optimal-Growth of Tropical Sea-Surface Temperature Anomalies. *Journal of Climate* 1995, **8**(8): 1999-2024.
25. Dias DF, Subramanian A, Zanna L, Miller AJ. Remote and local influences in forecasting Pacific SST: a linear inverse model and a multimodel ensemble study. *Climate Dynamics* 2019, **52**(5-6): 3183-3201.
26. Newman M, Sardeshmukh PD. Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures? *Geophysical Research Letters* 2017, **44**(16): 8520-8529.
27. Shin SI, Newman M. Seasonal Predictability of Global and North American Coastal Sea Surface Temperature and Height Anomalies. *Geophysical Research Letters* 2021, **48**(10).
28. Capotondi A, Sardeshmukh PD. Is El Niño really changing? *Geophysical Research Letters* 2017, **44**(16): 8548-8556.
29. Ault T, Deser C, Newman M, Emile-Geay J. Characterizing decadal to centennial variability in the equatorial Pacific during the last millennium. *Geophysical Research Letters* 2013, **40**(13): 3450-3456.
30. Newman M. An Empirical Benchmark for Decadal Forecasts of Global Surface Temperature Anomalies. *Journal of Climate* 2013, **26**(14): 5260-5269.

31. Frankignoul C, Gastineau G, Kwon YO. Estimation of the SST Response to Anthropogenic and External Forcing and Its Impact on the Atlantic Multidecadal Oscillation and the Pacific Decadal Oscillation. *Journal of Climate* 2017, **30**(24): 9871-9895.
32. Belkin IM. Rapid warming of Large Marine Ecosystems. *Progress in Oceanography* 2009, **81**(1-4): 207-213.
33. Capotondi A, Newman M, Xu T, Di Lorenzo E. An Optimal Precursor of Northeast Pacific Marine Heatwaves and Central Pacific El Niño events (under review). *Geophysical Research Letters* 2022.
34. Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, *et al.* Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research-Atmospheres* 2003, **108**(D14).
35. Balmaseda MA, Mogensen K, Weaver AT. Evaluation of the ECMWF ocean reanalysis system ORAS4. *Q J Roy Meteor Soc* 2013, **139**(674): 1132-1161.
36. Newman M, Alexander MA, Scott JD. An empirical model of tropical ocean dynamics. *Climate Dynamics* 2011, **37**(9-10): 1823-1841.
37. Zhao Y, Newman M, Capotondi A, Di Lorenzo E, Sun D. Removing the Effects of Tropical Dynamics from North Pacific Climate Variability. *Journal of Climate* 2021: 1-49.

605

606